

# Clustering Results

Zeotap Internship Assignment

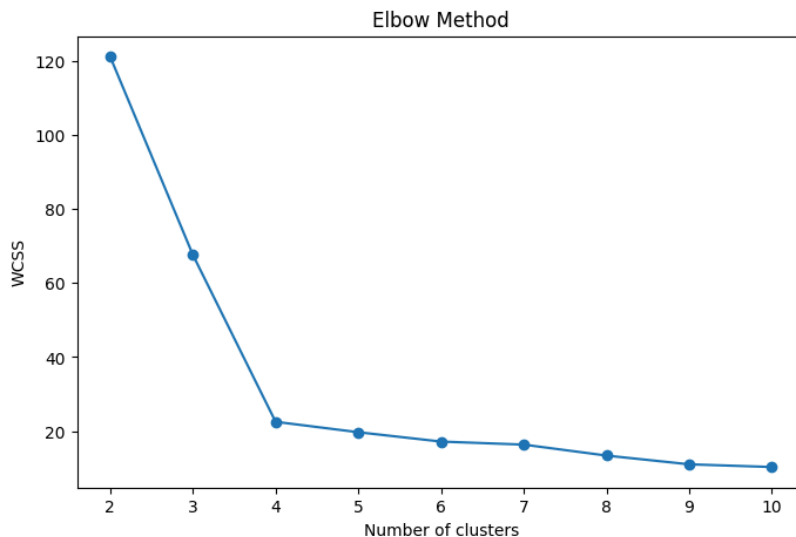
*By Ojas Sinha*

The following features were used for the clustering:

1. Total spending of customer
2. Average transaction value of customer
3. Number of transactions of customer
4. Region of customer (One-hot encoded)

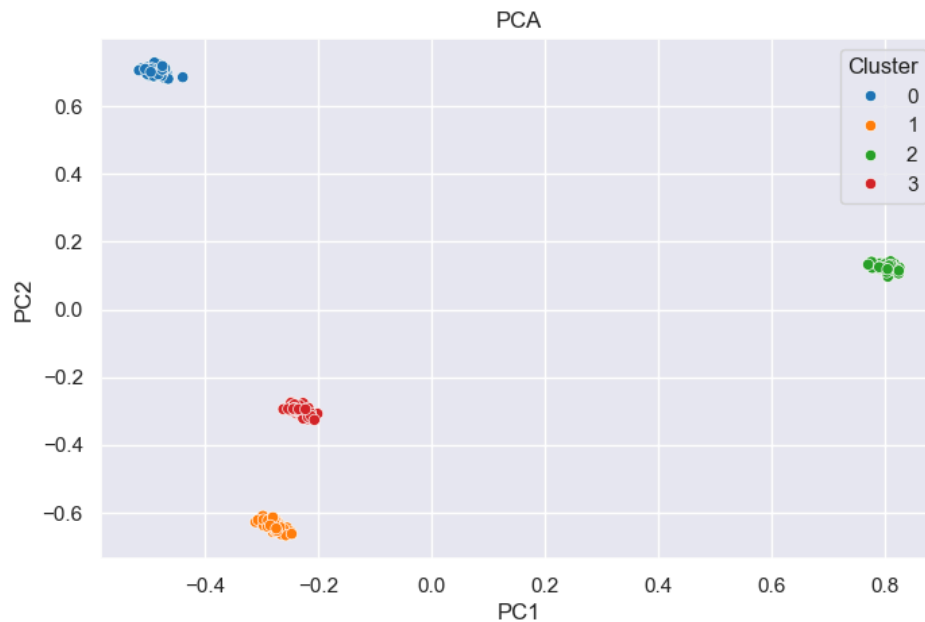
## Clustering Algorithm Used: K-Means Clustering

The **Elbow Method** was used to determine the optimal number of clusters. It involves plotting the sum of squared distances from each point to its assigned cluster center (within-cluster sum of squares) against the number of clusters. The idea is to identify the "elbow point" in the plot, where the rate of decrease sharply slows down, indicating diminishing returns for adding more clusters.



Here, we can see that the “elbow point” is at 4 clusters. Therefore, 4 is our optimal number of clusters.

## Results



*(Plotting clusters after decomposing into 2 axes using PCA)*

**Davies-Bouldin Index:** 0.4319307279812565

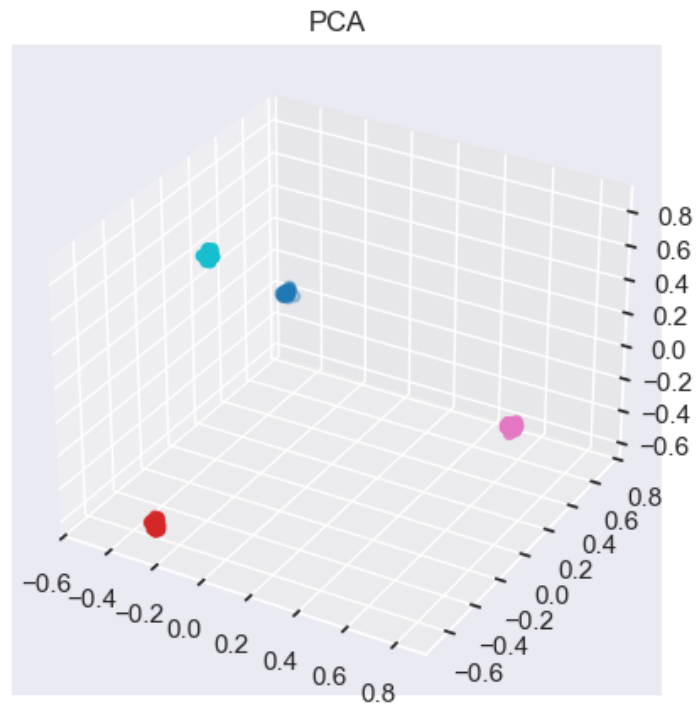
Such a low value in the DB index indicates that the clustering is of good quality. Lower values of DB index suggest that the clusters are compact and well-separated from each other.

**Silhouette Score:** 0.7167111219567973

Silhouette Scores close to 1 indicate that the sample is far away from the neighbouring clusters and very close to the cluster it is assigned to. Such a score suggests that our clustering is very good.

**Calinski-Harabasz Score:** 429.43176493484316

Generally higher Calinski-Harabasz scores indicate better-defined clusters, so a score of 429 suggests that our clusters are reasonably well-separated and compact.



*(Plotting clusters in 3D after decomposing into 3 axes using PCA)*